DNA Prediction in Craniofacial Appearance: Implications for the Identity Sciences

By A. Midori Albert and Charissa L. Wright

Abstract

This report relays findings from research published between 2010 and April 2015, pertinent to the expanding field of identity sciences, where DNA prediction of phenotypic traits related to craniofacial appearance, such as eye color, hair color, skin color, age and biogeographic origin, is the primary focus. Using the Cochrane method, over 40,000 articles were recognized as potentially allied to the research topic; of these, 475 were deemed most suitable, and consequently, analyzed. Results indicated 25 publications yielded information most germane to the research inquiry—how well can, from DNA, various craniofacial phenotypic traits be predicted, and what is the viability of applying this information to questions of validating or protecting individual identity. Discussed here is a review of the findings from the most recent and relevant studies. Results primarily show that eye color is promising for widespread use as a trait that can be reliably and accurately predicted from DNA. In need are improved methods of DNA extraction and analysis from the soft tissues as well as bones and teeth such that phenotypic trait prediction can become a workable and widespread tool in identity science casework.

1.0 Introduction

This report relays findings from research published between 2010 and April 2015, pertinent to the expanding field of identity sciences, where DNA prediction of phenotypic traits related to craniofacial appearance, such as eye color, hair color, skin color, age and biogeographic origin, is the primary focus. Discussed here is a review of the most recent and relevant studies pertaining to the use of DNA analyses to predict phenotypic characteristics for purposes of establishing identity. This review of the literature is meaningful to the identity sciences inasmuch as facial features are an important biometric in establishing positive identifications; and this work raises awareness concerning the scope and level of the technology presently available for predicting human craniofacial phenotypic variability.

2.0 Methods

The method utilized in this assessment was the Cochrane Criteria [1]. The Cochrane method encompasses developing a specified end goal, accessing and retrieving the data via search engines, and synthesizing the gathered information. This purpose-driven research, with designated guidelines, assists in reducing bias and supplying the best evidence for the most desirable outcome.

In this systematic review, key words and phrases were used to draw out the desired information from ScienceDirect, with ScienceDirect being used through the entirety of the search for consistent results. The decided process of data collection entailed systematically accumulating all of the relevant articles, mining the articles for information, categorizing the accumulated materials, and compiling the information. The results from the various searches are explained below.

3.0 Results

This review encompassed eight search iterations through ScienceDirect, with a total of 475 articles sifted through for the most relevant and up-to-date information (Table 1). The conducted search, as seen in Table 1, encompassed variations of "Human," "DNA/Genome," and "Prediction" as these concepts were the main focus of the research. Each iteration can be seen in the left-most column. The "Number of Articles Presented" was the total number of articles displayed when the search was first entered into ScienceDirect. "Number of Articles Searched Through" is the amount of the original number of articles presented that were examined as to their relevance to this study and was thus abandoned once it stopped providing useful information. If the article was deemed relevant, it was "Pulled" and those numbers can be seen in "Number of Articles

Pulled." The final column, "Number of Articles Used," displays the number of articles used in this current review. The key ideas and findings from the 25 articles extracted are presented in the next section.

Table 1: Literature Search.

Search Iteration	**Number of Articles Presented**	**Number of Articles Searched Through**	**Number of Articles Pulled**	**Number of Articles Used**
Human Genome Prediction Accuracy	7,970	125	12	11
Human DNA Prediction in Biometrics	565	50	1	0
Human DNA Predicting Hair Color	631	75	14	12
Human DNA Predicting Face Shape	2,498	50	0	0
Human Genome Craniofacial Predictions	483	50	0	0
Human Phenotypic Prediction from DNA	8,273	50	2	2
Phenotypic Genome Display	10,741	25	0	0
Human Phenotypic Genome Display	9,485	50	0	0
Totals:		**475**	**29**	**25**

All searches were completed on April 11, 2015 with the date constraint of 2010-present.

4.0 Discussion

4.1 DNA Prediction: Primarily Pigmentation

In the identity sciences, the ability to predict "externally visible characteristics" or EVCs [2], such as eye color, hair color, and skin color, is critical where DNA samples collected at crime scenes fail to result in matches with forensic database samples, and where there is the absence of eyewitnesses.

4.1.1 Single Nucleotide Polymorphisms

Kastelic and Drobnič [3] were among the earlier researchers to explore certain single nucleotide polymorphisms (SNPs) strongly related to pigment trait variability among European populations. Their study on a sample of Slovenian individuals resulted in an accuracy of 96% for eye and hair color using particular SNPs. The key to studies of this nature is determining which of a multitude of SNPs are the best predictors, and for which populations. Also relevant is the degree to which epistasis (where one gene may be dependent on one or more modifier genes) plays a role [4,5].

In New Zealand, current findings regarding the ability to accurately predict eye color are promising in terms of their potential use in forensic identification. Allwood and Harbison [2] constructed classification tree models using SNPs of known connection to eye color and tested the ability of these models to accurately predict eye color—blue vs. non-blue, brown vs. non-brown, and intermediate vs. non-intermediate—in unknown samples. A total of 19 SNPs from ten different genes known or believed to code for pigmentation were selected from previous research. The researchers developed four models for predicting eye color, the best of which resulted in 89% accuracy for blue and 94% accuracy for brown.

4.1.2 Introduction of IrisPlex

Chaitanya et al. [6] published their work on their DNA-based prediction system to test eye color on an international level (with 21 laboratories participating) where a sample of 1890 genotypes from diverse populations was examined. The system is known as the IrisPlex system; it comprises biological samples and a single forensically validated multiplex genotyping assay along with a statistically based model for prediction that was developed using genotypes and phenotypes from a sample of thousands of individuals. The researchers report that the IrisPlex system has an average of greater than 94% accuracy for predicting blue and brown eye color using six of the currently known best single nucleotide polymorphisms associated with eye color. This high level of accuracy and the ease with which IrisPlex can be applied demonstrates its potential widespread use in the forensic sciences in the very near future. However, it should be noted that while there is a high level of accuracy for predicting blue and brown eye colors, the intermediate eye colors are more challenging.

For example, Dembinksi and Picard [7] tested IrisPlex on a sample of North American individuals (i.e., US college campuses) and found 58% accuracy for brown eye color, 95% for blue eye color, and 11% for intermediate eye color. The relatively high levels of

population admixture resulting in extensive genotypic and phenotypic variability in the US may play a role in the efficacy of IrisPlex's use in the US.

4.1.3 Intermediate Eye colors and Admixture

Further, Ruiz et al. [8] addressed the issue of the challenge of eye color prediction for intermediate eye colors in their study of 416 individuals from six populations of north and south Europeans. Ruiz et al. [8] modified the genetic assays and statistical classification models; they confirmed a complex and continuous range of intermediate eye colors, distinct from blue and brown. The importance of this work is that it expands the capabilities of IrisPlex by yielding data on intermediate eye colors.

Indeed, Freire-Aradas et al. [9] have tested DNA prediction of eye color on samples of admixed individuals (those with European and South American ancestry) compared with non-admixed Europeans and found differences. Genome databases specifically addressing admixed populations are available and further study is underway.

Lima et al. [10] addressed the issue of predicting pigmentation in admixed populations. The authors establish that pigmentation is a variable and complex trait in humans, affected by the interplay of genes, biological factors (age, disease, hormones) and environmental factors (sun exposure) and (not mentioned) perhaps epigenetic phenomenon and that polymorphisms of these pigmentation genes have been correlated with a wide range of phenotypic appearances in eye, skin, and hair color in homogenous populations. Lima et al.'s [10] study focused on 483 individuals from a Brazilian sample who self-reported eye, skin, and hair coloring. The aim was to detect which genotypes were associated with which pigmentation phenotypes. Findings indicated certain different homozygous genotypes were strongly correlated with light skin, blue eyes, light hair, and darker traits, similar to previous studies of this nature.

With regard to the applicability of DNA phenotype (i.e., pigmentation in eye, hair, and skin color) prediction from human skeletal remains, Purps et al. [11] tested IrisPlex on a sample of individuals from Germany (n=102) and Turkey, East Asia, and Africa (n=81), then verified their results using a case study involving exhumed skeletal remains with degraded DNA. For the skeletal remains, IrisPlex predicted blue eyes and probable European descent. Purps et al. [11] support the future use of IrisPlex in forensic cases involving skeletal remains.

Moreover, Spichenok et al. [12] attest to the importance of using genetic information to predict EVCs (e.g., eye color, hair color, skin color) from decomposed or skeletonized human remains. Their validation study of 554 individuals using seven SNPs associated

with pigmentation supported the accuracy of eye color prediction, with skin color prediction being more challenging.

4.1.4 Introduction of HIrisPlex

Given the above studies that demonstrate the state of the art methodologies and techniques in the use of DNA predictions in forensic anthropology and the identification sciences, it is imperative that validation studies continue [13] since the application and widespread future use of these methods rests on their accuracy and replicability among diverse populations.

To this end, Walsh et al. [14] report that the recently introduced HIrisPlex assay (an extension of the earlier IrisPlex) meets the guidelines of the Scientific Working Group on DNA Analysis Methods (SWGDAM), a crucial prerequisite for its use in future forensic cases. They report the rise in efforts to employ DNA prediction of externally visible characteristics (EVCs), where DNA prediction is also known as Forensic DNA Phenotyping or 'DNA intelligence', for use in assisting law enforcement. Specifically, DNA prediction would aid in establishing the identity of unknown individuals by providing details of EVCs—beyond the biological profile determined skeletally in forensic anthropology (i.e., sex, ancestry, age at death, stature). DNA prediction would help establish the appearance of unknown suspects, offenders, and missing persons from biological samples such as blood, semen, saliva, and hair. Walsh et al. [14] explain in detail how HIrisPlex performs in determining human vs. nonhuman samples, developing complete DNA profiles, replicability across forensic laboratories, and using degraded DNA samples (in human remains up to several hundred years old). The authors note that the HIrisPlex system is also appropriate for analyzing old and ancient DNA in anthropological and evolutionary studies. See also Walsh et al. [15,16,17,18] for additional studies of IrisPlex and HIrisPlex.

Kayser [19] cautions that until specifics on individual appearances can be predicted accurately from DNA, conventional DNA profiling still needs to be performed. For more detailed biological information on the specifics of human pigmentation genetics, please see Liu et al. [20].

4.2 On the Horizon

4.2.1 Biogeographic Assessment

Aside from DNA prediction of human pigmentation traits, the ability to determine biogeographic ancestry is gaining interest. Gettings et al. [21] conducted a study in which they developed and tested a model to predict ancestry in the US population, based

on data available from forensic laboratories. Their model accurately assessed ancestry in 98.6% of the test set samples, and provided accurate eye color information in 61% of the samples of European ancestry tested (25% were inconclusive and 14% were incorrect). Gettings et al. [21] recommend their method for use in US forensic casework to provide additional information about an individual DNA sample when the Short Tandem Repeat (STR) method of DNA profiling has not resulted in an identification match.

4.2.2 Age

An important aspect of establishing the biological or identity profile in forensic anthropology is determining age at death in unknown individuals, typically accomplished through assessments of skeletal markers of growth and maturation as well as degenerative changes. The feasibility of using DNA to accurately and reliably assess age at death in unknown individuals is currently being explored. Zubakov et al. [22] note that earlier genetic methods of estimating age that rely on the accumulation of mitochondrial DNA deletions or on the shortening of telomeres (structures at the end of chromosomes that shorten each time DNA replicates during cell division) are problematic. Essentially, these methods result in low accuracies and inconsistencies related to technique; as a result, they are rendered inappropriate for forensic and identification science casework. In Zubakov et al.'s [22] study, the researchers examined the use of T-cell DNA extracted from blood as a method of estimating age. Their results yielded accurate and reliable predictions and the authors provide a method (PCR protocol) for applied use.

Saeed et al. [23] support the assertion that telomere length, while showing strong correlations with age also show a high level of inconsistency due to difficulties with technique and confounding variables that can produce unclear results. Telomere length and its association to age, however, remains a hopeful approach to estimating age from DNA and is presently the focus of myriad genetic studies [23].

DNA methylation (where methylation refers to a biochemical process where a methyl group is added to nucleotides consisting of either the base cytosine or adenine) is another area of genetic age estimation currently being explored. Yi et al. [24] conducted a study that determined eight gene fragments for which cytosine methylation showed a statistically significant correlation with age—in blood—in a sample of 40 individuals. Results indicated a high correlation between predicted age (based on a regression model) and actual age ($r=0.91$). The authors suggest that DNA methylation is likely to become a widely used biological marker for age estimation. They predict that analyses of particular sites on the genome could be routinely screened in forensic samples as a means to estimate age.

Although researchers are making strides with regard to using DNA to predict age, most of the positive results are derived from samples of DNA extracted from blood. In forensic anthropology, the ability to extract DNA from bones and teeth that is viable for use in estimating age—and any of the other EVCs (e.g., eye color, hair color, skin color)— is a goal, hopefully within reach in the very near future.

Zubakov et al. [22] note that biochemical methods, such as those based on the accumulation of D-aspartic acid, are often destructive to bones and teeth, and are limited by problems with technique and degradation of biological samples.

Beyond exploring the ability to estimate age using DNA, studies aimed at rendering a likeness of an individual's facial appearance based on genetic information is also underway.

4.2.3 Facial Composites

Claes et al. [25] have experimented with using 24 SNPs shown to influence facial variation. Overlaying these features on a "base-face" (averaged from sex and ancestry information) using an image blending technique, the accuracy of the predicted faces was tested using cross-validation. Although this approach to creating facial composites from DNA is preliminary, it shows where the science of DNA prediction and forensic anthropology are headed.

4.2.4 Public Health and Legislation

Another area of genetic research involving human variation and individualization concerns the ability to predict health and disease. Research of this nature has come to be known as "GRIPS" or Genetic Risk Prediction Studies. The relatively fast and ongoing process of uncovering genes coding for complex human diseases is a driving force behind the search for ways in which genetic risk models may be used in clinical health settings as well as in public health practices and legislation, particularly related to health insurance issues. Noteworthy is the article by Janssens et al. [26] in which the authors address the need for proper reporting of Genetic Risk Prediction Studies.

5.0 Conclusion

This review of DNA prediction of craniofacial morphological straits for use in the identity sciences speaks to the level of interaction evident among a multitude of disciplines answering challenging questions related to legal issues of human identity. Reporting on the state of research at the interface of these disciplines—anthropology and

molecular biology—serves to illuminate for all researchers and practitioners of identity science, whose foci, expertise, and applied interests vary widely, the most up to date happenings in at least this one specific area where legal concerns tied to human variation and individualization are at the crux.

REFERENCES

[1] Higgins JPT, Green S. Eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0, The Cochrane Collaboration, 2011. Available from:

[2] Allwood JS, Harbison S. SNP model development for the prediction of eye colour in New Zealand. Forensic Sci Int-Gen 2013; 7: 444-452.

[3] Kastelic V, Drobnič K. Single multiplex system of twelve SNPs: Validation and implementation for association of SNPs with human eye and hair color. Forensic Sci Int-Gen 2011; 3: e216-e217.

[4] Maroñas O, Phillips C, Söchtig J, Gomez-Tato A, Cruz R, Alvarez-Dios J, *et al*. Development of a forensic skin colour predictive test. Forensic Sci Int-Gen 2014; 13: 34-44.

[5] Pośpiech E, Wojas-Pelc A, Walsh S, Liu F, Maeda H, Ishikawa T, *et al.* The common occurrence of epistasis in the determination of uman pigmentation and its impact on DNA-based pigmentation phenotype prediction. Forensic Sci Int-Gen 2014; 11: 64-72.

[6] Chaitanya L, Walsh S, Andersen JD, Ansell R, Ballantyne K, Ballard D, *et al*. Collaborative EDNAP exercise on the IrisPlex system for DNA-based prediction of human eye colour. Forensic Sci Int-Gen 2014; 11: 241-251.

[7] Dembinski GM, Picard CJ. Evaluation of the IrisPlex DNA-based eye color prediction assay in a United States population. Forensic Sci Int-Gen 2014; 9: 111-117.

[8] Ruiz Y, Phillips C, Gomez-Tato A, Alvarez-Dios J, Casares de Cal M, Cruz R, *et al*. Further development of forensic eye color predictive tests. Forensic Sci Int-Gen 2013; 7: 28-40.

[9] Freire-Aradas A, Ruiz Y, Phillips C, Maroñas O, Söchtig J, Gómez Tato A, *et al*. Exploring iris colour prediction and ancestry inference in admixed population of South America. Forensic Sci Int-Gen 2014; 13: 3-9.

[10] Lima FdA, Gonçalves FdT, Fridman C. SLC24A5 and ASIP as phenotypic predictors in Brazilian population for forensic purposes. J Leg Med In Press 2015.

[11] Purps J, Geppert M, Nagy M, Roewer L. Evaluation of the IrisPlex eye colour prediction tool in a German Population sample. Forensic Sci Int-Gen 2011; 3: e202-e203.

[12] Spichenok O, Budimlija ZM, Mitchell AA, Jenny A, Kovacevic L, Marjanovic D, *et al*. Prediction of eye and skin color in diverse population using seven SNPs. Forensic Sci Int-Gen 2011; 5: 472-478.

[13] Pneuman A, Budimlija ZM, Caragine T, Prinz M, Wurmbach E. Verifcation of eye and skin color predictors in various populations. J Leg Med 2012; 14: 78-83.

[14] Walsh S, Chaitanya L, Clarisse L, Wirken L, Draus-Barini J, Kovatsi L, *et al*. Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. Forensic Sci Int-Gen 2014; 9: 150-161.

[15] Walsh S, Lindenbergh A, Zuniga SB, Sijen T, de Knijff P, Kayser M, *et al*. Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. Forensic Sci Int-Gen 2011a; 5: 464-471.

[16] Walsh S, Liu F, Ballantyne KN, van Oven M, Lao O, Kayser M. IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci Int-Gen 2011b; 5: 170-180.

[17] Walsh S, Liu F, Wollstein A, Kovatsi L, Ralf A, Kosiniak-Kamysz A, *et al*. The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. Forensic Sci Int-Gen 2013; 7: 98-115.

[18] Walsh S, Wollstein A, Liu F, Chakravarthy U, Rahu M, Seland JH, *et al*. DNA-based eye coulour prediction across Europe with the IrisPlex system. Forensic Sci Int-Gen 2012; 6: 330-340.

[19] Kayser M. Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes. Forensic Sci Int-Gen In Press 2015.

[20] Liu F, Wen B, Kayser M. Colorful DNA polymorphisms in humans. Sem Cell Dev Biol 2013; 24: 562-575.

[21] Gettings KB, Lai R, Johnson JL, Peck MA, Hart JA, Gordish-Dressman H, Schanfield MS, Podini DS. A 50-SNP assay for biogeographic ancestry and phenotype prediction in the U.S. population. Forensic Sci Int-Gen 2014; 8: 101-108.

[22] Zubakov D, Liu F, van Zelm MC, Vermeulen J, Oostra BA, van Duijn CM, *et al*. Estimating human age from T-cell DNA rearrangements. Curr Biol 2010; 20: R970-R971.

[23] Saeed M, Berlin RM, Cruz TD. Exploring the utility of genetic markers for predicting biological age. J Leg Med 2012; 14: 279-285.

[24] Yi SH, Xu LC, Mei K, Yang RZ, Huang DX. Isolation and identification of age-related DNA methylation markers for forensic age-prediction. Forensic Sci Int-Gen 2014; 1: 117-125.

[25] Claes P, Hill H, and Shriver MD. Toward DNA-based facial composites: Preliminary results and validation. Forensic Sci Int-Gen 2014; 13: 208-216.

[26] Janssens ACJW, Ioannidis JPA, Bedrosian S, Boffetta P, Dolan SM, Dowling N, *et al*. Strengthening the reporting of Genetic Risk Prediction Studies (GRIPS): explanation and elaboration. J Clin Epidemiol 2011; 64: e1-e22.